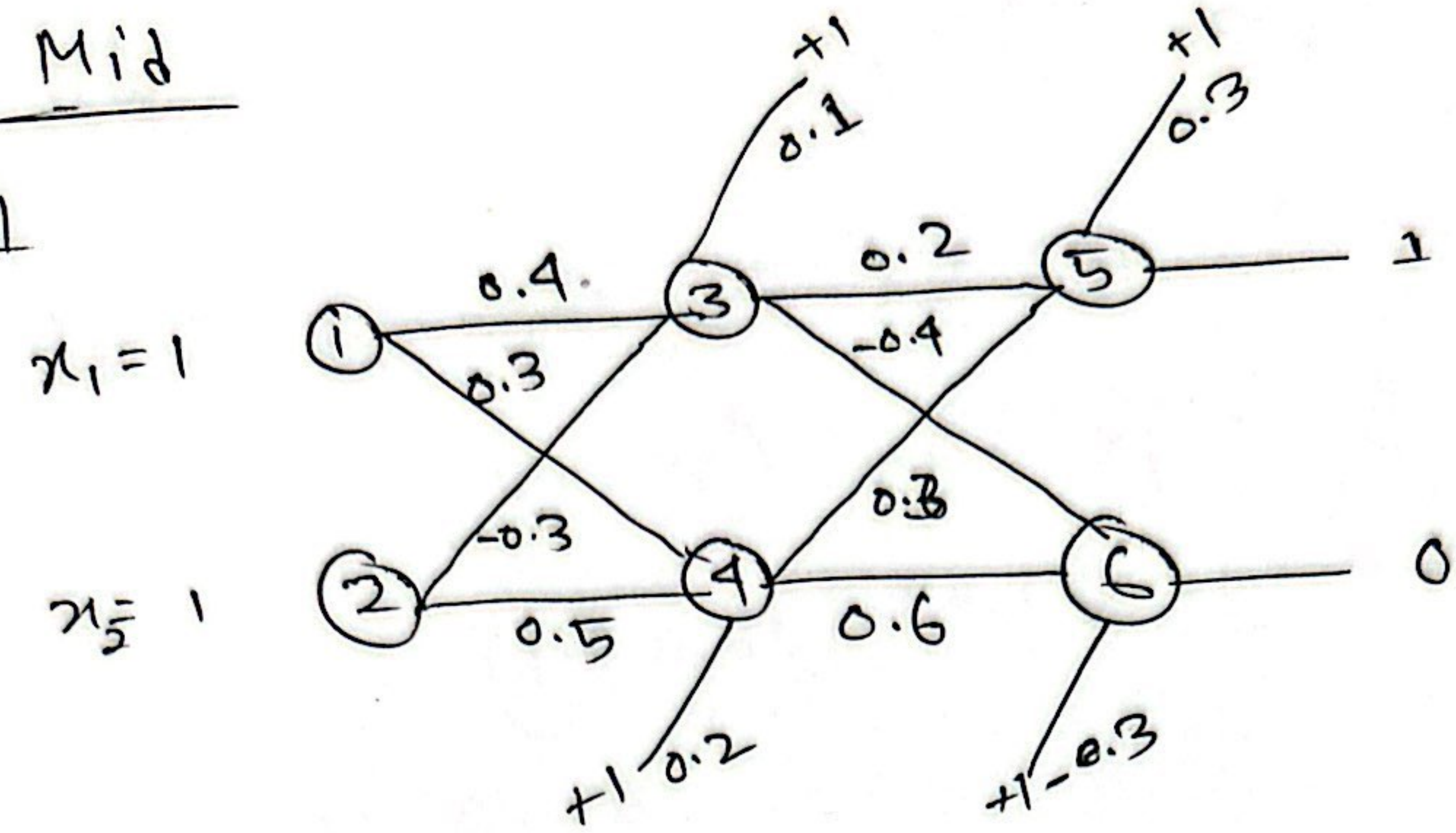
Mid

8)



$I_j \quad \Sigma(\text{input}_j \times w_j)$

$I_3 = x_1 \times W_{13} + x_2 \times W_{23} + 1 \times W_{03}$

$= 1 \times 0.4 + 1 \times (-0.3) + 1 \times 0.1$

$= 0.2$

$I_4 = x_1 \times W_{14} + x_2 \times W_{24} + 1 \times W_{04}$

$= 1 \times 0.3 + 1 \times 0.5 + 1 \times 0.2$

$= 1$

$I_5 = O_3 \times W_{35} + O_4 \times W_{45} + 1 \times W_{05}$

$= 0.55 \times 0.2 + 0.73 \times 0.3 + 1 \times 0.3$

$= 0.629$

$I_6 = O_3 \times W_{36} + O_4 \times W_{46} + 1 \times W_{06}$

$= 0.55 \times (-0.4) + 0.73 \times 0.6 + 1 \times (-0.3)$

$= -0.082$

O_j ~~(input)~~, $\frac{1}{1+e^{-I_j}}$

$O_3 = \frac{1}{1+e^{-0.2}}$

$= 0.55$

$O_4 = \frac{1}{1+e^{-1}}$

$= 0.73$

$O_5 = \frac{1}{1+e^{-0.629}}$

$= 0.65$

$O_6 = \frac{1}{1+e^{-(-0.082)}}$

$= 0.48$

$$\Delta_j = \text{Error}_j \times O_j \times (1 - O_j)$$

$$\Delta_5 = (\text{target} - O_j) \times O_j \times (1 - O_j)$$

$$= (1 - O_5) \times O_5 \times (1 - O_5)$$

$$= (1 - 0.65) \times 0.65 \times (1 - 0.65)$$

$$= 0.08$$

$$\Delta_6 = (0 - 0.48) \times 0.48 \times (\cancel{1} - \cancel{0}.48)$$

$$= -0.12$$

$$\Delta_3 = \text{Error}_3 \times O_3 \times (1 - O_3)$$

$$= (\Delta_5 \times W_{35} + \Delta_6 \times W_{36}) \times O_3 \times (1 - O_3)$$

$$= \{0.08 \times 0.2 + (-0.12) \times (-0.4)\} \times 0.55 \times (1 - 0.55)$$

$$= 0.016$$

$$\Delta_4 = \text{Error}_4 \times O_4 \times (1 - O_3)$$

$$= (\Delta_5 \times W_{45} + \Delta_6 \times W_{46}) \times O_4 \times (1 - O_4)$$

$$= \{0.08 \times 0.3 + (-0.12) \times 0.6\} \times 0.73 \times (1 - 0.73)$$

$$= -0.0095$$

change the weight of every links

$$W_{ij} = W_{ij}(old) + \alpha \times O_i \times \Delta_6$$

$$W_{13} = W_{13}(old) + \alpha \times O_1 \times \Delta_3$$

$$= 0.4 + 0.9 \times 1 \times 0.016$$

$$= 0.414$$

$$W_{14} = W_{14}(old) + \alpha \times O_1 \times \Delta 4$$

$$= 0.3 + 0.9 \times 1 \times (-0.0095)$$

$$= 0.299$$

$$W_{23} = -0.3 + 0.9 \times 1 \times 0.016$$

$$= -0.29$$

$$W_{24} = 0.5 + 0.9 \times 1 \times (-0.0095)$$

$$= 0.49$$

$$W_{35} = 0.2 + 0.9 \times 0.55 \times (0.08)$$

$$= 0.24$$

$$W_{36} = -0.4 + 0.9 \times 0.55 \times (-0.12)$$

$$= -0.46$$

$$w_{45} = 0.3 + 0.9 \times 0.73 \times 0.08$$

$$= 0.35$$

$$w_{46} = 0.6 + 0.9 \times 0.73 \times (-0.12)$$

$$= 0.52$$

3) To avoid local minima.

~~1) Starting~~

1) Random Restart (Hill climbing):

~~it~~ it can be started from a random position and update the weight and fix the error. The process to be continued.

2) Batch mode: After itirating all the example and sumation of the error it should update the weight and fix it,

3) Pattern mode: After itirating every single example it fixes the error and update the weight.

For momentum term increase

$$\Delta W_{jk}(new) = \alpha \, \Delta W_{jk}(old) + \eta \, a_j \, \Delta_k$$

$$W_{jk}(new) = W_{jk}(old) + \Delta W_{jk}(new)$$

$$\therefore \Delta W_{jk}(t) = \alpha \, \Delta W_{jk}(t-1) + \eta \, g(t)$$

$$\Delta W_{jk}(1) = \alpha \, \Delta W_{jk}(0) + \eta \, g(1)$$

$$\Delta W_{jk}(2) = \alpha \, \Delta W_{jk}(1) + \eta \, g(2)$$

$$= \alpha \left(\alpha \, \Delta W_{jk}(0) + \eta \, g(1) \right) + \eta \, g(2)$$

$$= \alpha^2 \Delta W_{jk}(0) + \alpha \eta \, g(1) + \eta \, g(2)$$

$$\Delta W_{jk}(3) = \alpha \, \Delta W_{jk}(2) + \eta \, g(3)$$

$$= \alpha \left(\alpha^2 \Delta W_{jk}(0) + \alpha \eta \, g(1) + \eta \, g(2) \right) + \eta \, g(3)$$

$$= \alpha^3 \Delta W_{jk}(0) + \alpha^2 \eta \, g(1) + \alpha \eta \, g(2) + \eta \, g(3)$$

$$= \alpha^3 \Delta W_{jk}(0) + \eta \left(\alpha^2 g(1) + \alpha^1 n(2) + \alpha^0 g(3) \right)$$

$$\Delta w_{jk}(T) = \alpha^{T} \Delta w_{jk}(0) + \eta\left(\alpha^{T-1} g(1) + \alpha^{T-2} g(2) + \cdots g(T)\right)$$

$$\Delta w_{jk}(T) = \underbrace{\alpha^{T} \Delta w_{jk}(0)}_{0} + \eta \sum_{t=0}^{T-1} \alpha^{t} g(T-t)$$

$$= \eta\left(1+\alpha+\alpha^{2}+ \cdots +\alpha^{T-1}\right) g$$

$$= \eta \frac{1}{1-\alpha} \times g$$

$$= \frac{\eta}{1-\alpha} \times g$$ [where g is gradient which is very small]

So, momentum could be increase by increasing the value of η.

As $\Delta w_{jk}(T) \propto \eta$

And η is the controller of the contribution of the momentum term to the update direction.

41

Cross entropy function:

- The cross-entropy function is a commonly used cost function in machine learning, particularly in classification tasks. It measures the difference between the predicted probability distribution and the true probability distribution of the target variable.

- And it also a well formed function which doesn't trap at local minima.

$$T = \sum_{i=1}^{N} \sum_{k=1}^{k} \left(y_k^{(i)} \ln \hat{y}_k^{(i)} + \left(1 - y_k^{(i)}\right) \ln\left(1 - \hat{y}_k^{(i)}\right) \right)$$

T is the cross entropy cost function

where $y_k^{(i)}$ is actual output

$\hat{y}_k^{(i)}$ is predicted output

minimum T if $y_k(i) = \hat{y}_k(i)$

maximum T if $y_k(i) \neq \hat{y}_k(i)$ that means completely different

The activation function could be sigmoid function for the output node.

Because: Segmoid function maps the output of the last layer of the neural network to a value between 0 and 1.

5) control Size of Neural Network :

Cost function Regularization

$$J = \sum_{i=1}^{N} \mathcal{E}(i) + \alpha \mathcal{E}_p(\underline{w})$$

Here the size minimize by J and it depends on $\mathcal{E}_p(\underline{w})$ which is regularization factor.

$$\mathcal{E}_p(\underline{w}) = \sum_{k=1}^{k} b(w_k)^2$$

$$= \sum_{k=1}^{k} \frac{w_k^2}{w_0^2 + w_k^2}$$

if $\frac{w_k^2}{w_0^2 + w_k^2} <$ threshold

then this weighted example delete from ~~set~~ network.